Statistica I modulo - CdS SIAFA *(2023/24)*
Esame 30 gennaio 2024

Cognome e nome ______________________________ **Matricola** ____________

1. Si consideri la seguente distribuzione di frequenze in classi del numero di ore giornaliere trascorse sull'app *Tennis is nice!* da 336 tennisti non professonisti:

Classe	n_i
0-1	88
1-2	162
2-3	66
3-4	18
4-8	2
TOT	336

(a) Calcolare la media aritmetica.

Per la media aritmetica si calcolano i valori centrali di ogni classe sui quali si calcola una media aritmetica ponderata con pesi pari alle frequenze assolute:

$$\bar{x} = \frac{0,5 \cdot 88 + 1,5 \cdot 162 + \ldots + 6 \cdot 2}{336} = \frac{527}{336} = 1,568$$

(b) Calcolare il 10° percentile, spiegandone il significato.

Per il calcolo del 10° percentile, si calcolano le frequenze relative cumulate F_i:

Classe	n_i	f_i	F_i
0-1	88	0,262	0,262
1-2	162	0,482	0,744
2-3	66	0,196	0,940
3-4	18	0,054	0,994
4-8	2	0,006	1
TOT	336	1	

Si imposta quindi la proporzione:

$$(0,262 - 0) : (1 - 0) = (0,10 - 0) : (P_{10} - 0)$$

da cui

$$P_{10} = \frac{1 \cdot 0,10}{0,262} = 0,38$$

(c) Rappresentare l'istogramma.

Dopo aver calcolato le densità di frequenza ($h_i = n_i/a_i$), si ottiene:

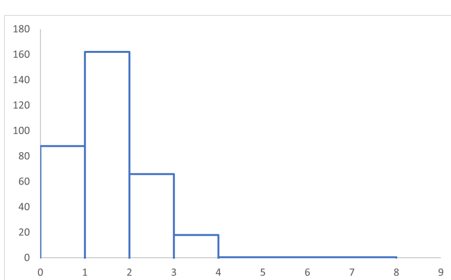

Figura 1: Istogramma.

2. Per 6 tennisti si registrano per il numero di ore trascorse sull'app *Tennis is nice!* nella settimana precedente il torneo *Città di Napoli* (X) e il numero di partite vinte nel torneo (Y):

Numero ore (X)	Vittorie (Y)
1,5	4
2,5	3
2	4
1	1
4,4	5
1,5	1

(a) Rappresentare il diagramma di dispersione.

Il diagramma di dispersione è riportato nella Figura 2.

(b) Identificare i coefficienti della regressione (con Y variabile dipendente) e rappresentare la retta di regressione.

Per individuare la retta di regressione si svolgono i calcoli sotto riportati dopo aver individuato $\bar{x} = 2,15$ *e* $\bar{y} = 3$.

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	Prodotto	$(x_i - \bar{x})^2$
1,5	4	-0,65	1	-0,65	0,4225
2,5	3	0,35	0	0	0,1225
2	4	-0,15	1	-0,15	0,0225
1	1	-1,15	-2	2,3	1,3225
4,4	5	2,25	2	4,5	5,0625
1,5	1	-0,65	-2	1,3	0,4225
TOT				7,3	7,375

Dunque

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{6}[(x_i - \bar{x})(y_i - \bar{y})]}{\sum_{i=1}^{6}(x_i - \bar{x})^2} = \frac{7,3}{7,375} = 0,99$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 3 - 0,99 \cdot 2,15 = 0,87$$

L'equazione della retta di regressione è dunque

$$\hat{Y} = 0,87 + 0,99X$$

La retta di regressione è riportata nella Figura 2.

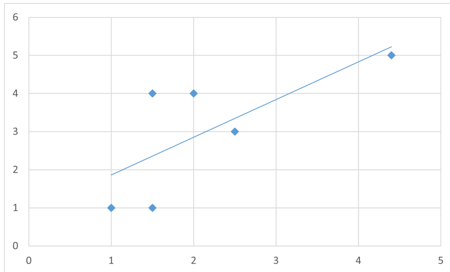

Figura 2: Diagramma di dispersione e retta di regressione.

(c) Prevedere il numero di vittorie se $X = 3$.

Con $X = 3$, il numero previsto di vittorie è

$$\hat{Y}(3) = 0,87 + 0,99 \cdot 3 = 3,84$$

3. Il numero di sponsor per un torneo alla prima edizione segue una variabile casuale di Poisson con $\lambda = 3,4$.

   (a) Fornire un'interpretazione di λ.

   Il parametro λ rappresenta il numero medio di successi nell'arco temporale specificato.

   (b) Qual è la probabilità che vi siano 5 sponsor?

   La probabilità che vi siano 5 sponsor è

$$P(X = 5) = \frac{e^{-3,4} 3,4^5}{5!} = 0,126$$

   (c) Qual è la probabilità che vi siano almeno 2 sponsor?

   La probabilità che vi siano almeno 2 sponsor è

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - 0,033 - 0,113 = 0,853$$

   (d) Considerando che il numero di sponsor in un torneo è indipendente dal numero di sponsor in altri tornei, qual è la probabilità che in 2 tornei vi siano più di 2 sponsor?

   Definita la variabile casuale $Y = X + X$, si ha che $Y \sim Po(6,8)$. La probabilità che in 2 tornei vi siano più di 2 sponsor è

$$P(Y > 2) = 1 - P(Y = 0) - P(Y = 1) - P(Y = 2) = 1 - 0,001 - 0,008 - 0.026 = 0,966$$